WHAT IS CLAIMED IS:

1. A method of downloading data sets by a plurality of web crawlers from among a plurality of host computers, comprising the steps of:

assigning a web crawler identifier to each one of the plurality of web crawlers;

for each respective web crawler:

downloading at least one data set that includes addresses of one or more referred data sets;

identifying the addresses of the one or more referred data sets, wherein each identified address includes a host computer identifier;

for each identified address:

generating a representation of the host computer identifier;

determining a web crawler identifier to which the representation corresponds; and

when the determined web crawler identifier is not assigned to the respective web crawler, sending the identified address to the web crawler to which the determined web crawler identifier is assigned.

2. The method of claim 1, wherein

the plurality of web crawlers consists of n web crawlers; and

generating the representation includes computing a function of the host computer identifier to generate an integer value that is a member of a set of n predefined distinct values.

3. The method of claim 1, wherein

the plurality of web crawlers consists of n web crawlers; and

generating the representation includes computing a hash function of the host computer identifier to generate an intermediate value V, and computing V modulo n.

4. The method of claim 1, wherein the sending step includes:

determining a web crawler address for the web crawler to which the determined web crawler identifier is assigned;

transmitting the identified data set address to the destination web crawler at the determined web crawler address.

5. A method of downloading data sets by a plurality of web crawlers from among a plurality of host computers, comprising the steps of:

for each respective web crawler:

receiving addresses of one or more data sets from each of the plurality of web crawlers other than the respective web crawler;

for each received address:

determining if the address has been previously stored; and

if this determination is negative, storing the address.

6. A web crawler system for downloading data set addresses from among a plurality of host computers, comprising:

a plurality of web crawlers, wherein each web crawler has been assigned a web crawler identifier;

for each respective web crawler:

a main web crawler module for downloading and processing data sets stored on a plurality of host computers, the main web crawler module identifying addresses of the one or more referred data sets in the downloaded data sets, wherein each identified address includes a host computer identifier; and

an address distribution module for processing the identified addresses, the address distribution module including instructions for:

generating a representation of the host computer identifier, wherein the representation corresponds to one of the web crawler identifiers;

determining a web crawler identifier to which the representation corresponds; and

when the determined web crawler identifier is not assigned to the respective web crawler, sending the identified address to a destination web crawler comprising the web crawler to which the determined web crawler identifier is assigned.

7. The web crawler system of claim 6 wherein

the plurality of web crawlers consists of n web crawlers; and

the address distribution module's instructions for generating the representation includes instructions for computing a hash function of the host computer identifier to generate an intermediate value V, and computing V modulo n.

8. The web crawler system of claim 6, further comprising:

for each respective web crawler, a web crawler interface for transmitting the identified address to the destination web crawler and for receiving identified addresses from each of the plurality of web crawlers other than the respective web crawler.

9. The web crawler system of claim 6, further comprising:

for each respective web crawler, a lookup table storing for each of the plurality of web crawler identifiers a corresponding web crawler address, said lookup table for use by the address distribution module in determining a web crawler address to which to send the identified data set address.

10. A computer program product for use in conjunction with a web crawler system wherein each web crawler is assigned a web crawler identifier, the computer program product comprising a computer readable storage medium and a computer program mechanism embedded therein, the computer program mechanism comprising:

a main web crawler module for downloading and processing data sets stored on a plurality of host computers, the main web crawler module identifying addresses of the one or more referred data sets in the downloaded data sets, wherein each identified address includes a host computer identifier; and

an address distribution module for processing the identified addresses, the address distribution module including instructions for:

generating a representation of the host computer identifier, wherein the representation corresponds to one of the web crawler identifiers;

determining a web crawler identifier to which the representation corresponds; and

when the determined web crawler identifier is not assigned to the respective web crawler, sending the identified address to a destination web crawler comprising the web crawler to which the determined web crawler identifier is assigned.

11. The computer program product of claim 10, wherein:

the web crawler system consists of n web crawlers; and

the address distribution module's instructions for generating the representation includes instructions for computing a function of the host computer identifier to generate an integer value that is a member of a set of n predefined distinct values.

12. The computer program product of claim 10, wherein:

the web crawler system consists of n web crawlers; and

the address distribution module's instructions for generating the representation includes instructions for computing a hash function of the host computer identifier to generate an intermediate value V, and computing V modulo n.

13. The computer program product of claim 10, further comprising:

a web crawler interface for transmitting the identified address to the destination web crawler and for receiving identified addresses from each of the plurality of web crawlers other than the respective web crawler.

14. The computer program product of claim 10, further comprising:

a lookup table storing for each of the plurality of web crawler identifiers a corresponding web crawler address, said lookup table for use by the address distribution module in determining a web crawler address to which to send the identified data set address.